

SOUTHERN FOREST EXPERIMENT STATION

E. L. Demmon, Director

New Orleans, La.



A METHOD FOR INVENTORYING PLANTING STOCK
IN FOREST NURSERIES

by

R. A. Chapman,
Associate Silviculturist,
Southern Forest Experiment Station.

The Occasional Papers of the Southern Forest Experiment Station present information on current Southern forestry problems under investigation at the Station. In some cases these contributions were first presented as addresses to a limited group of people, and as "occasional papers" they can reach a much wider audience. In other cases, they are summaries of investigations prepared especially to give a report of the progress made in a particular field of research. In any case, the statements herein contained should be considered subject to correction or modification as further data are obtained.

Note: Assistance in the preparation of these materials was furnished by the personnel of Works Progress Administration Official Project 365-64-3-7.

A METHOD FOR INVENTORYING PLANTING STOCK
IN FOREST NURSERIES

By R. A. Chapman,
Associate Silviculturist,^{1/}
Southern Forest Experiment Station

Each year forest nurserymen are faced with the problem of estimating the number of seedlings they can ship from their nurseries. Since errors in the estimates may seriously affect the planting programs, during the past 2 years the Southern Forest Experiment Station has been attempting to improve nursery-inventory procedure. The inventory method that appears most promising consists of making two estimates: one of the total number of seedlings in the entire nursery, or a natural subdivision of it; and another of the percentage of plantable seedlings. With the data from these two estimates, the number of plantable seedlings in the nursery (or subdivision) can be calculated readily.

Estimating Total Number of Seedlings

To estimate the total number of seedlings, it is necessary to use sampling units. These may be of various sizes, but at this Station a unit 1 ft. wide by 4 ft. long has been used. The length of the sampling unit (4 ft.) corresponds to the width of the standard Forest Service seedbed, and by taking a sample that extends entirely across the bed, variation due to border effect is eliminated from the sampling variation. The width of the sample might be somewhat greater, but it should be remembered that for a given total area in samples, it is better to have many small samples than a few large ones.

The size of these sampling units must be measured accurately. As an aid in improving the accuracy, the Forest Service nurseries in the South use a light, rigid, steel frame 1 ft. by 4 ft. in size. This frame is placed across the seedbed, and seedlings falling inside it are recorded as being in the sample. When counting the seedlings in the sample, care should be exercised to record all the seedlings; counting all culls, regardless of size, eliminates any variations in personal judgment concerning what seedlings to count and what to omit, and also avoids any bias in grading.

The number of samples to take for any inventory depends on the variability of the beds being sampled and the accuracy desired. It has been observed that the large variation in numbers of seedlings per sampling unit is due mainly to variation between nursery beds. (A nursery bed is here considered as a unit 100 ft. long and 4 ft. wide.) Wherever there is a large variation between beds, the accuracy of the estimate of number of plantable seedlings can be increased by sampling in such a way that an equal number of sampling units falls in each bed. For longleaf and slash pine seedbeds in large nurseries, it has been found that if 1 percent of each bed is included in the samples, the error of the mean is about 3 percent.

^{1/} Acknowledgement is hereby made of the useful suggestions received from P. C. Wakeley during the course of this study.

When only one sample is taken in each bed, however, it will be impossible to compute the accuracy of the estimate. If it is desired to compute the error of the mean, then at least two randomly located samples must be taken in each bed. A mechanical system of sampling which can not be considered as random is often used. It is possible (and even probable) that this type of sampling will give as accurate results as randomly selected plots, but since the plots are spaced mechanically, it is impossible to obtain a valid estimate of the accuracy of the inventory. The accuracy of mechanical sampling depends on the skill of the nurseryman in arranging his grid work of samples so that all sources of variability are sampled.

The procedure for obtaining a random sample can be shown most easily by an illustration. Assume that it is desired to take a 2-percent sample for inventory purposes. Also assume that the beds being sampled are 100 ft. long and the sampling unit to be used is 1 ft. wide. Thus each bed will contain 100 possible sampling units. The numbers from 1 to 100 are recorded separately on small slips of paper or tags that easily can be mixed thoroughly. Two numbers are drawn out and noted. These two numbers indicate the position of the sampling units to be counted in the first bed. The two numbers are put back in the box, the numbers are mixed again, and two more numbers are drawn. These numbers represent the sampling units to count for the second bed. This procedure is repeated for all beds. It should be noted that it is essential to mix the numbers thoroughly after each draw.^{2/}

Having determined which samples are to be counted, it is necessary to find these sampling units in the field. If the man making the count is a fairly accurate pacer, the samples can be located by pacing. Assume that sample 75 is drawn; then the observer will pace in 75 ft. from the end of the bed and place the frame on the bed at this point. To avoid a possible bias, the observer should place the frame where his pace shows it should be, regardless of the kind of sample it will give. Always pace in from the same end of the bed.

Estimating the Percentage of Plantable Seedlings

The estimate of the percentage of plantable seedlings is usually obtained by digging the seedlings on a number of sampling units, but such a procedure is very laborious. It may happen also that an individual nurseryman "has a good eye" for estimating the percentage of plantable seedlings. Such an estimate is permissible, but it should be remembered that it is very difficult to make an unbiased estimate.

Also for some species of trees, the root and top characters or sizes

^{2/} If one has access to a table of random numbers, it will not be necessary to record the sampling number on slips or tags and draw out samples; instead one can read the random numbers directly from the tables. Two sets of random numbers can be found in "Tracts for computers, No. 15. Random sampling numbers," by L. H. C. Tippett, Cambridge Univ. Press, London, 1927; and in "Statistical tables for biological, agricultural, and medical research," by R. A. Fisher and F. Yates, published by Oliver and Boyd, London, England, 1938.

are very highly correlated, in which case top grade can be used as a substitute for seedling grade. Where this is not true, and if the nurseryman lacks unusual skill in grading by tops alone, then the seedlings on a number of sampling units will have to be dug and graded. These units should be the same size as the units used to estimate the total number of seedlings. Care should be exercised to make sure that all seedlings are recorded that would have been if the samples had been used to estimate the total number of seedlings.

The average percentage of plantable seedlings could be estimated directly by digging all the sampling units used to estimate the total number of seedlings, but since this would be very expensive, use is made of the relation of percentage of seedlings plantable to total number of seedlings.

By making use of this relation it is possible to reduce the number of samples that have to be dug, which can be selected to cover the range of total number of seedlings per sampling unit. In choosing these samples for digging, care should be taken to let nothing but the range in density influence the choice, and particularly not to pick samples because of any observed or suspected plantable percentage. It is not necessary that the samples used in making the counts of total number of seedlings be used as part of this percent-plantable sample. If the range in density is not very large, 30 or 40 samples will be sufficient for most work, but the more samples taken the greater will be the accuracy.

The main inventory, which usually is made a month or so before actual lifting begins, is taken to supply an estimate of the number of plantable seedlings that will be present at lifting time. If the sampling is done when the seedlings are growing, later there is likely to be a change in grade, and it will be necessary to apply some sort of correction factor. This can be determined by experience or by noting, for each sample dug, which seedlings that are now classed as culls will be plantable by the time lifting begins.

In estimating change of grade, an endeavor should be made to estimate all causes of change. Two important sources of error in grade change are seedling growth, which causes an increase, and mechanical injury, which causes a decrease, in the number of plantable seedlings. The first of these errors depends on the ability of the grader to estimate what change will take place under the expected weather conditions. The second of these errors depends on soil conditions, on the lifting equipment, and on the care in its operation. If the seedlings are lifted when the soil is hard and tends to break into large lumps when the lifter is used, then a number of the seedlings may have to be culled because their roots have been broken off. If the lifting equipment is not skillfully used, a number of the seedlings will be injured and have to be culled. If one is either too optimistic or too pessimistic and assumes too large or too small a correction factor, an over- or underestimate will result.

Although one might think that the sample for determining total number of seedlings should be corrected for mortality during the period between inventory and lifting, this is not necessary if the mortality of plantable seedlings is estimated on the samples that are lifted.

All of these suggested corrections may make it appear that the estimate finally arrived at is only a guess. This criticism is partly true, but it should be noted that any procedure for estimating the plantable seedlings at a future date is subject to the same criticism. The accuracy of these corrections can be improved by additional study, and each nurseryman should try to develop a set of correction factors for his own nursery.

Calculation of Total Number of Plantable Seedlings

The samples taken to determine the total number of seedlings should be summed and averaged, in order to supply an estimate of the average total number of seedlings for all sampling units in the nursery or subdivision. From the other set of samples, on which number of plantable seedlings has been estimated, it is possible to compute the percentage of the total number of seedlings plantable. This percentage should be computed for each sample and plotted on a sheet of cross-section paper, plotting percentage of plantable seedlings over total number of seedlings. For the data available on southern pines, this curve has been found to be in most cases a nearly horizontal straight line passing through the mean percentage plantable.

Since a linear relation is indicated, the curve should pass through the mean of the percent plantable and the mean of the total number of seedlings. From this curve of percentage of plantable seedlings over total number of seedlings can be read the percentage of plantable seedlings corresponding to the average total number of seedlings obtained from the first set of samples.

The average number of plantable seedlings per sampling unit is the product of: (1) the average number of seedlings per sampling unit; (2) the corrected average percentage of plantable seedlings estimated from the relationship obtained from the second set of samples; and (3) a constant ($\frac{1}{100}$) to convert the percentages to actual numbers.

The total number of plantable seedlings is then obtained by multiplying the average number of plantable seedlings per sampling unit by the total number of units being sampled.

To illustrate this computation, let us assume that the sampling has been completed and the following information compiled:

1. From the density samples, it has been found that the mean number of seedlings per sampling unit is 140.

2. From figure 1A, which represents the relation of percentage of plantable seedlings to density, the percentage of plantable seedlings corresponding to a density of 140 has been found to be 89.0.

3. The average number of plantable seedlings per sampling unit is then

$$140 \times \frac{89.0}{100} = 124.60$$

4. If the subdivision of the nursery contains 200 seedbeds, each

containing 100 sampling units, then the total number of seedlings is

$$200 \times 100 \times 124.60 = 2,492,000$$

Summary of Procedure

The procedure for inventorying seedlings in forest nurseries (just described in detail) can be outlined briefly as follows:

1. Obtain from one set of samples an estimate of the average number of all seedlings per sampling unit (a convenient unit is 1 ft. x 4 ft.).
2. Obtain from another set of samples the relation of percentage of plantable seedlings per sampling unit to the total number of seedlings per unit, and express by means of a curve.
3. From the curve obtained in step 2, determine the percentage of plantable seedlings that corresponds to the average number of seedlings found in step 1.
4. Multiply the percentage of plantable seedlings found in step 3 by the average number of seedlings found in step 1, and divide by 100 to get the estimated number of plantable seedlings per sampling unit.
5. Multiply this estimated number of plantable seedlings per sampling unit by the total number of sampling units in the group of beds.

(APPENDIX)

The Fitting of Curves

The steps discussed in the body of this report seem logical and simple, but when the nurseryman undertakes to apply them in his own nursery, he may find that the drawing of the curve of plantable percentage over the total number of seedlings (as called for in step 2) is an obstacle to his application of the inventory method. This obstacle is easily overcome, however, since it arises largely from unfamiliarity with the process of curve fitting.

Fitting curves to the type of data ordinarily obtained in nursery inventories involves no great difficulties, as can be shown by the example in table 1, in which columns 2 and 3 contain a record of the number of plantable and culled seedlings in each of twenty sample plots. The total number of seedlings in the sample is, of course, the sum of the two columns, as recorded in column 4. Column 5 contains the percentage of plantable seedlings—column 2 divided by column 4. A plot of the data in columns 4 and 5 is shown in figure 1A.

Table 1.— Original and derived data required for fitting rectilinear curves by the method of least squares

Observation number	Number of plantable seedlings	Number of culled seedlings	Total number of seedlings X	Percentage of plant- able seed- lings Y	X^2	XY
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	57	7	64	89.1	4,096	5,702.4
2	59	10	69	85.5	4,761	5,899.5
3	62	3	65	95.4	4,225	6,201.0
4	81	7	88	92.0	7,744	8,096.0
5	88	13	101	87.1	10,201	8,797.1
6	95	9	104	91.3	10,816	9,495.2
7	109	10	119	91.6	14,161	10,900.4
8	112	18	130	86.2	16,900	11,206.0
9	117	11	128	91.4	16,384	11,699.2
10	120	10	130	92.3	16,900	11,999.0
11	150	12	162	92.6	26,244	15,001.2
12	146	19	165	88.5	27,225	14,602.5
13	165	19	184	89.7	33,856	16,504.8
14	157	38	195	80.5	38,025	15,697.5
15	186	13	199	93.5	39,601	18,606.5
16	175	36	211	82.9	44,521	17,491.9
17	204	26	230	88.7	52,900	20,401.0
18	230	15	245	93.9	60,025	23,005.5
19	206	52	258	79.8	66,564	20,588.4
20	<u>198</u>	<u>73</u>	<u>271</u>	<u>73.1</u>	<u>73,441</u>	<u>19,810.1</u>
Total	2,717	401	3,118	1,765.1	568,590	271,705.2

There are two ways that a curve can be fitted to the data in the figure: one is by a mathematical method known as the "method of least squares;" the other is a graphic procedure commonly used by foresters. The mathematical method of curve fitting is more accurate (and therefore preferable) when the relationship between the two factors (percentage of plantable seedlings and total number of seedlings) can be represented as a straight line, but to use this method to the best advantage requires some knowledge of equations and curve shapes; the choice of the wrong equation may introduce as much error as careless curve fitting, especially if one endeavors to extrapolate the curve beyond the main range of the data. When the relationship between the two variables is curved, the graphic procedure may be easier, but great care should be exercised in drawing the curve. Draw the simplest, smoothest curve possible, and do not let the end points unduly influence its shape.

The Method of Least Squares

An inspection of figure 1A indicates that the relationship between the two factors there plotted can be represented by a straight line. The general equation for a straight line is

$$Y = a + b (X)$$

where

$\underline{Y}$ is the dependent variable (in this problem, the percentage of plantable seedlings); $\underline{X}$ is the independent variable (i.e., the total number of seedlings per sample); and $\underline{a}$ and $\underline{b}$ are constants whose values can be obtained by solving the two equations

$$a = \frac{[\Sigma Y] [\Sigma X^2] - [\Sigma X] [\Sigma XY]}{N [\Sigma X^2] - [\Sigma X] [\Sigma X]} \quad (1)$$

$$b = \frac{N [\Sigma XY] - [\Sigma X] [\Sigma Y]}{N [\Sigma X^2] - [\Sigma X] [\Sigma X]} \quad (2)$$

In these equations, and in this example,

N = the number of observations, = 20;

ΣX = the sum of all the $\underline{X}$ values, or the total of column 4, = 3,118;

ΣX^2 = the sum of the squares of all the $\underline{X}$ values, or the total of column 6, = 568,590;

ΣY = the sum of all the $\underline{Y}$ values, or the total of column 5, = 1,765.1;

and ΣXY = the sum of the products of the individual $\underline{X}$ and $\underline{Y}$, or the total of column 7, = 271,705.2.

Then, substituting the above values in equations (1) and (2), we get

$$a = 94.82$$

$$b = -.0421$$

The equation for the straight line shown in figure 1A thus becomes

$$Y = 94.82 - .0421 (X)$$

The Graphic Method

If the data are not numerous, the curve can be fitted graphically to the individual points, but if the data are numerous the procedure becomes very tedious. To reduce the work of curve fitting, the data can be grouped by classes of the independent variable (in this case, the total number of seedlings per sample) and averaged. The curve can then be fitted to the averages of the various classes. The object of grouping is to reduce the number of individual points without losing any of the pertinent information supplied by the sample. If the data are distributed fairly uniformly over the range of the independent variable, about 10 groups or classes form a convenient and workable number. The data from table 1 have been grouped and assembled in table 2, columns 1 to 6.

The number of observations in each class-interval or group is shown in column 2. The sums of the $\bar{X}$ and $\bar{Y}$ values for each class-interval are shown in columns 3 and 4. These sums divided by the frequency values in column 2 are given in columns 5 and 6, respectively, and these means (columns 5 and 6) are the points plotted in figure 1B.

After plotting the averages, an attempt should be made to locate the best-fitting curve, which, if the relationship can be represented as a straight line, should pass through the mean $\bar{X}$ and $\bar{Y}$ (shown as a cross in fig. 1B). The slope of the line, which must be determined by inspection, should be that for which the sum of the squared residuals is a minimum. When computing the sum of the squared residuals for grouped data, it should be noted that each squared residual must be multiplied by the class frequency before summing.

To illustrate this curve-fitting procedure in detail, consider curve I, figure 1B. The curved value of $\bar{Y}$ for each $\bar{X}$ listed in column 5 (table 2) is read off from curve I and recorded in column 7. The residual Z listed in column 8 is the actual $\bar{Y}$ (column 6) minus the estimated $\bar{Y}$ (column 7); i.e., $-2.1 = 90.0 - 92.1$. In columns 9 and 10 are listed the products of the frequency column 2 and the residual column 8. If the product is positive, the result is entered in column 9; if negative, it is entered in column 10. The sums of these two columns, 20.6 and 20.4, are nearly equal, indicating that the curved values have been read very accurately. To obtain the sum of the squared residuals, the values in columns 9 or 10 are multiplied by the value in column 8 and entered in column 11, the sum of which (174.70) is the sum of the squared residuals.

To judge whether any other straight line fits the data better, consider the arbitrarily drawn lines, II and III, figure 1B. For these two lines, computations similar to those in columns 7 to 11 give the sums of the squared residuals 185.82 and 192.29, respectively. Since 174.70, the sum of the squared residuals for curve I, is smaller than either 185.82 or 192.29, curve I is the best fit.

Table 2.- Summary of data in table 1 and derived data required for fitting curves by the graphic method

Class interval	Frequency = $\underline{f}$	Sum of $\underline{X}$	Sum of $\underline{Y}$	Mean $\underline{X}$	Mean $\underline{Y}$	Line #1				
						Estimates $\underline{Y}'$	Residual $\underline{Y} - \underline{Y}' = \underline{Z}$	$\underline{f} \times (+\underline{Z})$	$\underline{f} \times (-\underline{Z})$	$\underline{f} \times \underline{Z}^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
60 - 79	3	198	270.0	66.0	90.0	92.1	-2.1		6.3	13.23
80 - 99	1	88	92.0	88.0	92.0	91.2	.8	.8		.64
100 - 119	3	324	270.0	108.0	90.0	90.3	-.3		.9	.27
120 - 139	3	388	269.9	129.3	90.0	89.4	.6	1.8		1.08
140 - 159										
160 - 179	2	327	181.1	163.5	90.6	87.9	2.7	5.4		14.58
180 - 199	3	578	263.7	192.7	87.9	86.7	1.2	3.6		4.32
200 - 219	1	211	82.9	211.0	82.9	85.9	-3.0		3.0	9.00
220 - 239	1	230	88.7	230.0	88.7	85.1	3.6	3.6		12.96
240 - 259	2	503	173.7	251.5	86.8	84.1	2.7	5.4		14.58
260 - 279	1	271	73.1	271.0	73.1	83.3	-10.2		10.2	104.04
Total	20	3,118	1,765.1	155.9	88.3			20.6	20.4	174.70

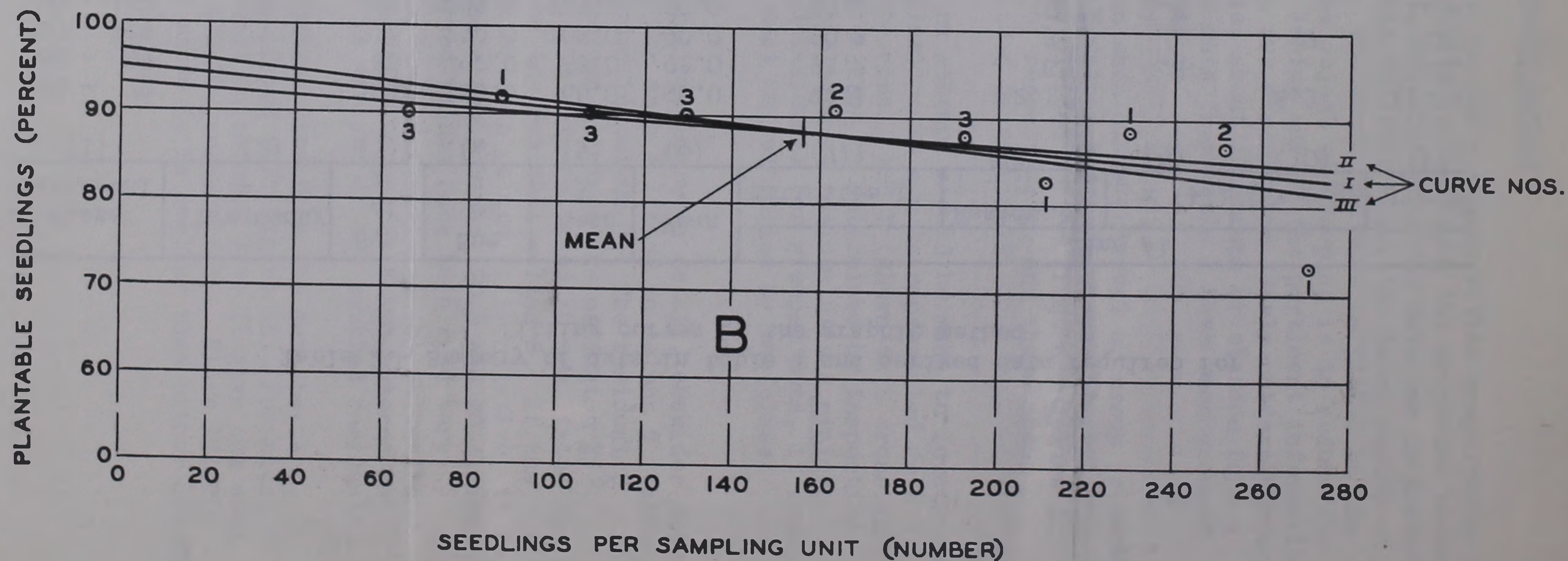
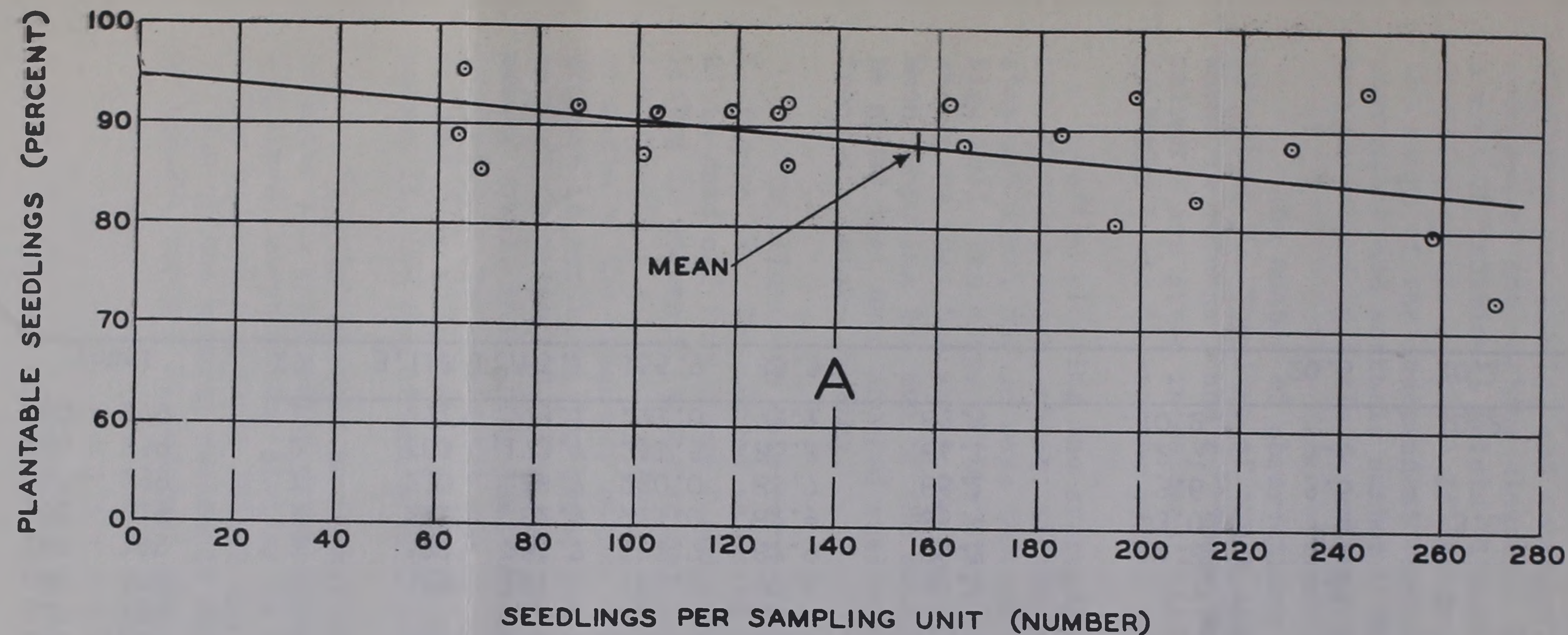


FIGURE 1 - SEEDLINGS PER SAMPLING UNIT PLOTTED AGAINST PERCENTAGE OF PLANTABLE SEEDLINGS.